# Supplementary materials for "Probabilistic reconstruction of truncated particle trajectories on a closed surface"

*Yunjiao Lu*[1,2,$], *Pierre Hodara*[1,$], *Charles Kervrann*[2], *Alain Trubuil*[1,*]

[1]*INRA, UR 1404, MaIAGE, Université Paris-Saclay, Jouy-en-Josas, France*
[2]*Inria, Centre de Recherche Bretagne-Atlantique, EPC SERPICO, Rennes, France*

[$]*these authors contributed equally to this work*
[*]*corresponding author*

## 1 WHEN THE LENGTH OF MOVIES EQUALS TO 2.5 MIN, THE ESTIMATION ERRORS OF $\hat{\tau}_\alpha$ AND $\hat{\tau}_D$ CAUSE THE FAILURE OF THE CONNECTION PROCEDURE

### 1.1 *Estimation of birth rate $\tau_\alpha$ and of death rate $\tau_d$*

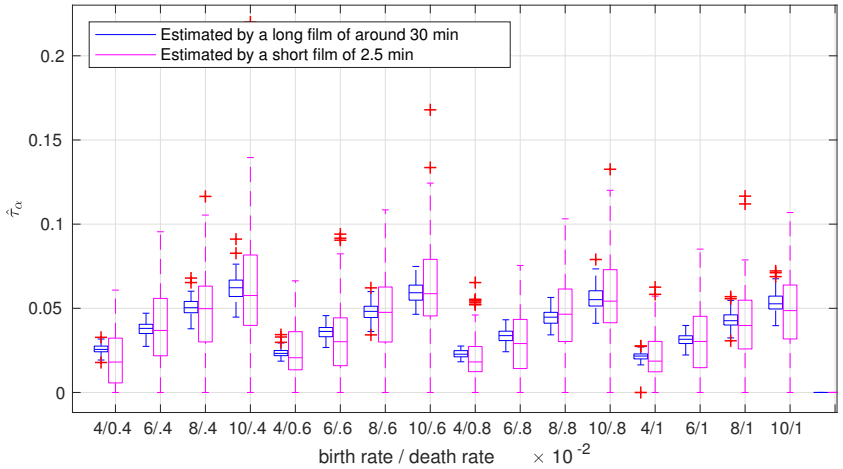


Figure 1: The estimation of arrival rate $\tau_\alpha$ with different $\lambda$ and $\tau_d$. Magenta boxes represent $\hat{\tau}_\alpha$ estimated by 2.5-min movies and blue boxes by 30-min movies.

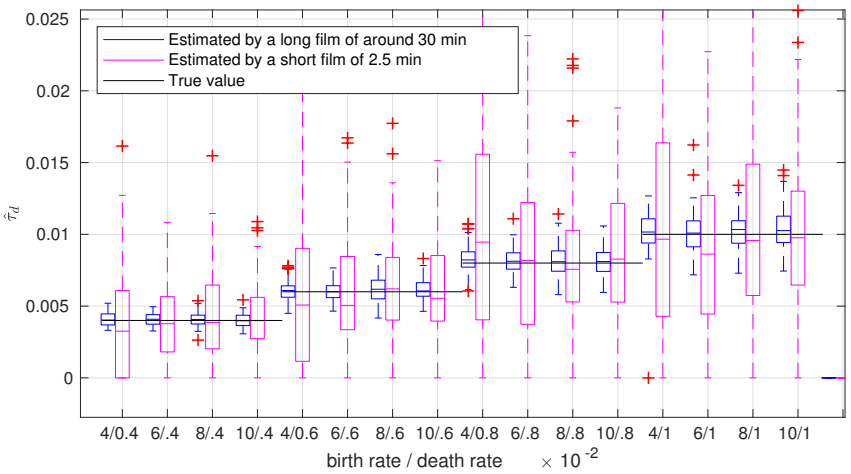


Figure 2: The estimation of death rate $\tau_d$ with different $\lambda$ and $\tau_d$. Magenta boxes represent $\hat{\tau}_d$ estimated by 2.5-min movies and blue boxes by 30-min movies. Black horizontal lines represent the true value of $\tau_d$.

The arrival rate $\tau_\alpha$ is estimated from 2.5 min movies and the result is presented in Fig. 1. Compared to Fig. 8 where 5-min movies are used, the estimations with 2.5-min movies have bigger variance. With 2.5-min movies, a considerable proportion of the estimations equal to zero. Similar conclusion can be made for $\hat{\tau}_d$ comparing Fig. 9 and Fig. 2.

It is shown that both in Figs 1 and 2, the estimators $\hat{\tau}_\alpha$ and $\hat{\tau}_d$ have many zero values. This phenomenon will deny the birth and death of particles, and force the connection of observed tracklets. The result of connection is presented in the next section and it shows that these biases of parameter estimation have severe influence on the tracklets connection performance.

### 1.2 *Evaluation of the connection procedure*

Figs. 3 and 4 show the results of tracklets connection measured by ARI, according to different settings of birth rate $\lambda$ and death rate $\tau_d$. To be specified, $\tilde{\tau}_\alpha$ and $\tilde{\tau}_d$ are estimators when $T_S = 30$ min, on the contrary, the estimators obtained with 2.5-min movies are denoted as $\hat{\tau}_\alpha$ and $\hat{\tau}_d$. Concerning the ex-

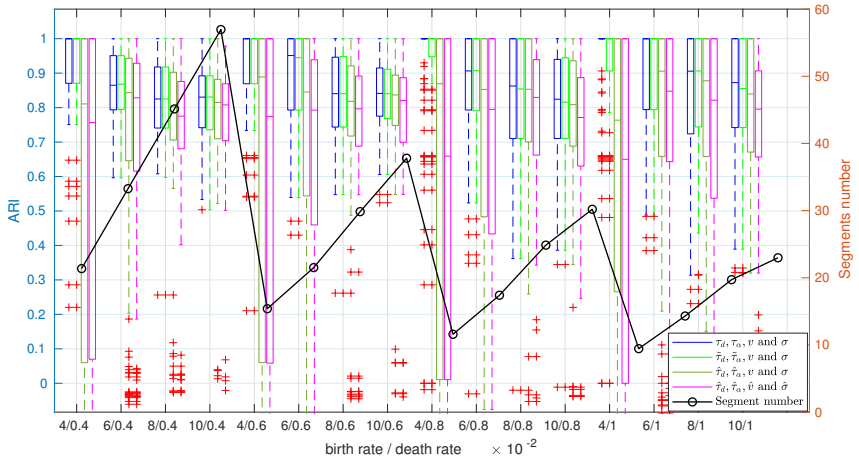


Figure 3: Connection performance comparison. Blue (resp. green, dark green and magenta) boxes represent ARI values obtained with parameters $(\tau_d, \tau_\alpha, \mathbf{v}, \sigma)$ (resp. $(\tilde{\tau}_d, \tilde{\tau}_\alpha, \mathbf{v}, \sigma)$, $(\hat{\tau}_d, \hat{\tau}_\alpha, \mathbf{v}, \sigma)$ and $(\hat{\tau}_d, \hat{\tau}_\alpha, \hat{\mathbf{v}}, \hat{\sigma})$). Light green boxes, representing the result when $\tilde{\tau}_\alpha, \tilde{\tau}_d$ (estimators with 30-min movies) and true value $v, \sigma$ are used, show performance as good as the blue boxes, where true parameters values are used. However, the dark green boxes, where $\hat{\tau}_\alpha, \hat{\tau}_d$ (with 2.5-min movies) and $v, \sigma$ are used, show much degraded results.

periment with true parameters (blue boxes in both figures), for all settings of $\lambda$ and $\tau_d$, the connection results are satisfying. However, for the experiments with estimated parameters (magenta boxes in both figures), we see clearly the failure of the connection represented by boxes with very large variance and low ARI values.

To identify which estimators are the main cause of this failure, many intermediate experiments are designed. To remind, $\tilde{\tau}_\alpha$ and $\tilde{\tau}_d$ represent the estimators of $\tau_\alpha$ and $\tau_d$ from 30-min movies, while $\hat{\tau}_d$ and $\hat{\tau}_\alpha$ from 2.5-min movies. Through Fig. 3 and 4, it can be concluded that the error of estimators $\hat{\tau}_d, \hat{\tau}_\alpha$ is the main cause of the dramatic decrease of ARI when all the estimators $\hat{\tau}_d, \hat{\tau}_\alpha, \hat{v}$ and $\hat{\sigma}$ are used.

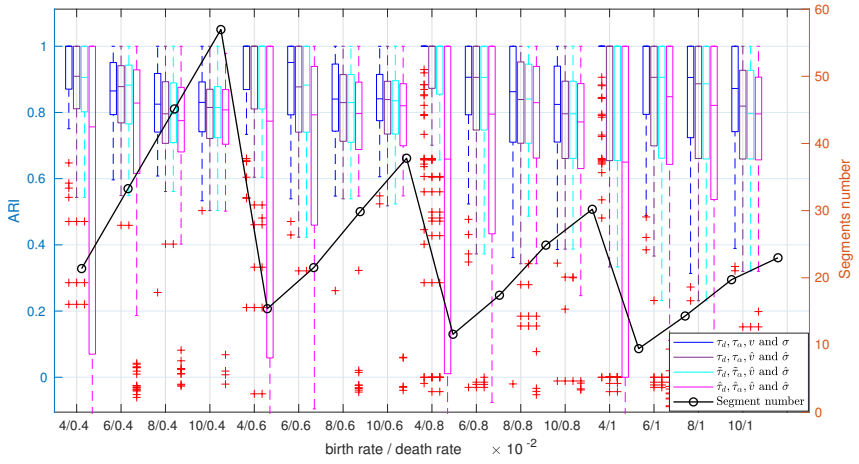


Figure 4: Connection performance comparison. Blue (resp. violet, cyan and magenta) boxes represent ARI values obtained with parameters $(\tau_d, \tau_\alpha, \mathbf{v}, \sigma)$ (resp. $(\tau_d, \tau_\alpha, \hat{\mathbf{v}}, \hat{\sigma})$, $(\tilde{\tau}_d, \tilde{\tau}_\alpha, \hat{\mathbf{v}}, \hat{\sigma})$, and $(\hat{\tau}_d, \hat{\tau}_\alpha, \hat{\mathbf{v}}, \hat{\sigma})$). Boxes with blue, violet, and cyan colors are similar, which means that the estimators $\hat{v}, \hat{\sigma}, \tilde{\tau}_\alpha$ and $\tilde{\tau}_d$ do not cause degradation of the connection. Only when $\hat{\tau}_\alpha$ and $\hat{\tau}_d$ are used, shown in magenta boxes, the connection results degrades.

The accuracy of $\hat{\tau}_\alpha$ and of $\hat{\tau}_d$ increases as the total observed time $T_S$ increases (see Fig. 7). Therefore, in the study, we chose movies of 5 min to estimate $\tau_\alpha$ and $\tau_d$ and then to evaluate the connection performance.

## 2 ANALYSIS OF ERRORS

In this section, we evaluated the connection error caused by randomness. We display in Fig. 5 the scatter plots of ARI value vs $K(c^t) - K(c^*)$, where $c^*$ denotes the optimal configuration calculated by the "Tracklets Connection Algorithm" while $c^t$ is the true configuration. Each scatter plot displays the results of 100 simulations for a given combination of $\lambda$ and $\tau_d$.

The difference between $K(c^t)$ and $K(c^*)$ is always positive or null, showing the optimization procedure works correctly to find the optimal solution. When $K(c^t) > K(c^*)$, it means that the configuration $c^*$ has a higher prob-

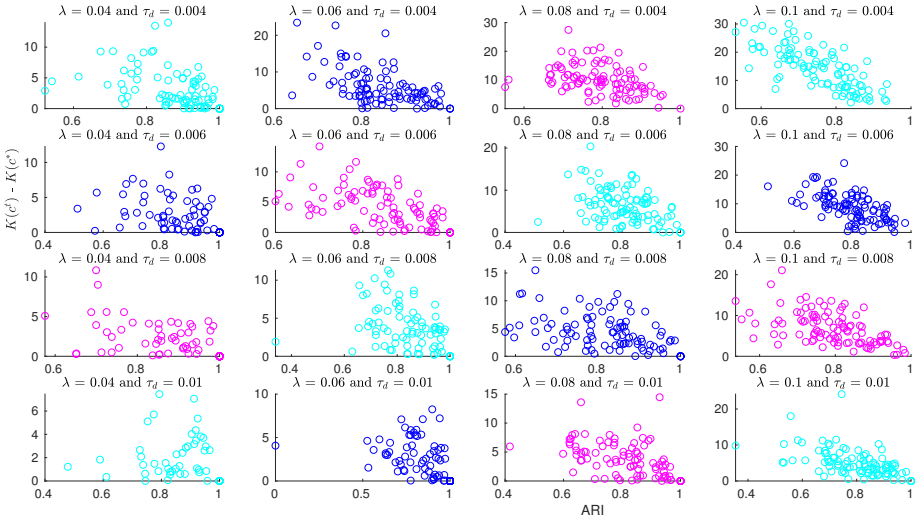


Figure 5: The difference between $K(c^t)$ and $K(c^*)$ versus ARI for different values of birth rate and death rate.

ability than the true realization $c^t$, which can be due to randomness. We can notice that ARI decreases as soon as $K(c^t) - K(c^*)$ increases. This error occurs when the realization is significantly different from the optimal configuration. Overall, ARI values are generally above 0.7.

Overall, we observe an increment of ARI when the difference decreases. The point clouds are diagonally shaped from top left to bottom right, showing a continuity that bigger the difference between $K(c^t)$ and $K(c^*)$, lower are the values of ARI.

Therefore, to improve the performance of the connection model, it is needed to find some characteristics of different trajectories to distinguish the true realization. To investigate the connection errors, we display in Fig. 5 the scatter plots of ARI value vs. the difference between $K(c^t)$ and $K(c^*)$, where $c^*$ denotes the optimal configuration calculated by the "tracklets Connection Algorithm" while $c^t$ is the true configuration. Each scatter plot is associated with a given combination of $\lambda$ and $\tau_d$. We considered 100 replications for

this experiment. On the left column, the connection is perfect in most case. Many points overlapped and located at coordinates (1,0). Overall, we observe an increment of ARI when the difference decreases. This can be due to low values of ARI obtained when the ground truth does not correspond to the optimal configuration.

## 3 THE BOUNDARY OF $P(C)$

With the optimization algorithm (Eq. 15), we note $c_i$ the i$^{\text{th}}$ optimal solution. Note that $N_c$ is the number of all possible configurations given $S$. For any $1 \leq n \leq N_c$, we have

$$\sum_{i=1}^{n} Q(c_i) \leq \sum_{i=1}^{N_c} Q(c_i) \leq \sum_{i=1}^{n} Q(c_i) + (N_c - n)Q(c_n)$$

Using Eq. 11, it gives

$$\frac{Q(c)}{\sum_{i=1}^{n} Q(c_i) + (\tilde{N}_c - n)Q(c_n)} \leq P(c) \leq \frac{Q(c)}{\sum_{i=1}^{n} Q(c_i)},$$

where $\tilde{N}_c$ is the number of possible configurations calculated through the number of inputs and outputs. We know that $\tilde{N}_c$ is bigger than $N_c$ because some configurations counted in $\tilde{N}_c$ are not compatible according to the time of inputs and outputs. Unfortunately we don't know the exact $N_c$.

Note $l(c,n) = \frac{Q(c)}{\sum_{i=1}^{n} Q(c_i) + (\tilde{N}_c - n)Q(c_n)}$ and $u(c,n) = \frac{Q(c)}{\sum_{i=1}^{n} Q(c_i)}$. In order to guarantee the precision of the probability, we choose $n^*, 1 \leq n^* \leq N_c$, big enough to satisfy, for $\alpha > 0$,

$$\frac{(\tilde{N}_c - n^*)Q(c_{n^*})}{\sum_{i=1}^{n^*} Q(c_i)} < \alpha. \tag{1}$$

This gives us

$$\frac{u(c,n^*)}{l(c,n^*)} = \frac{\sum_{i=1}^{n^*} Q(c_i) + (\tilde{N}_c - n^*)Q(c_{n^*})}{\sum_{i=1}^{n^*} Q(c_i)} = 1 + \frac{(\tilde{N}_c - n^*)Q(c_{n^*})}{\sum_{i=1}^{n^*} Q(c_i)} < 1 + \alpha$$

which ensures that the upper and lower bounds are close from each other. As a result, it gives

$$\sum_{i=1}^{n^*} P(c_i) > \sum_{i=1}^{n^*} l(c_i, n^*) > \frac{1}{1+\alpha}.$$

In other words, this means that the set of configurations $c_i$ up to $n^*$ correspond to an highly likely event for $\frac{1}{1+\alpha}$ close to 1.

## 4 SUMMARY OF NOTATIONS USEFUL FOR THE EVALUATION OF THE LIKELIHOOD

- $H$: the length of the cylinder,
- $L$: the length of the circumference of the cylinder,
- $l$: the length of the part of the of circumference that is observed,
- $l_u$: the length of the part of the of circumference that is not observed,
- $l_e$: the difference between the length of the unobserved region and the observed region,
- $B_c$: for a given configuration (reconstruction) of trajectories $c$, the subset of trajectories born in the unobserved region and seen on the border $\{-l\} \times [0, H]$,
- $D_c$: for a given configuration (reconstruction) of trajectories $c$, the subset of trajectories seen on the border $\{0\} \times [0, H]$ and died in the unobserved region,
- $\Delta t$: time stepsize between two consecutive observations,
- $N_l$: the number of particles born in the observed region and reaching the border $\{0\} \times [0, H]$,
- $p_x$: the probability of birth of a particle in a strip of width $x$, to the left side of the border $\{-l\} \times [0, H]$,

- $\hat{p}_x$: an estimator of $p_x$,
- $S$: the observation set of all trajectories,
- $S_o$: the set of tracklets having an output in $\{0\} \times [0, H]$,
- $S_l^*$: the set of tracklets having an input in $\{-l\} \times [0, H]$ and an ouput in $\{0\} \times [0, H]$, that is crossing the observed region,
- $S_r$: sample of points inside a restricted region inside the observed region. This region should allow to decide if a particle died or is just moving outside the observed region.
- $\tau_\alpha$: arrival rate at border $\{l\} \times [0, H]$ of particles born in the unobserved region $]-L, l[$,
- $\hat{\tau}_\alpha$: estimator of the arrival rate at border $\{l\} \times [0, H]$,
- $\tau_d$: death rate of particles,
- $\hat{\tau}_d$: estimator of the death rate,
- $T_S$: time duration of observation.